A quantitative approach for modelling the influence of currency of information on decision-making under uncertainty

Bernd Heinrich[a;*] and Diana Hristova[a; 1]

a: *Department of Management Information Systems, University of Regensburg, Universitätsstraße. 31, 93053 Regensburg, Germany*

*: Corresponding author: Bernd.Heinrich@wiwi.uni-regensburg.de (B. Heinrich)

Phone: +49 941 643 6100, Fax: +49 941 943 6120

1: Diana.Hristova@wiwi.uni-regensburg.de (D. Hristova)

Phone: +49 941 643 6104, Fax: +49 941 943 6120

Acknowledgement:

The Version of Record of this manuscript has been published and is available in Journal of Decision Systems, 23 Sep 2015,

http://www.tandfonline.com/doi/abs/10.1080/12460125.2015.1080494

A quantitative approach for modelling the influence of currency of information on decision-making under uncertainty

Stored information, used to support decision-making, can be outdated. Existing metrics for currency provide an indication about the correspondence between this stored information and its real-world counterpart. In the case of low currency, this information cannot be effectively used to support decision-making, although the decision-maker can probably learn from it. Our first objective is to develop an extended metric referring to currency, which provides an indication about the real-world information at the time of measurement, based on the stored information. Thus, the decision can be adjusted and the value of the stored information increased. Therefore, as a second objective, we propose a quantitative approach for modelling the influence of currency on decision-making by extending the normative concept of the value of information. Finally, we demonstrate the relevance of our approach by applying it to two real-world scenarios from the field of sales management in customer relationship management.

Keywords: information quality; decision-making; metric for currency; value of information; customer relationship management

1. Introduction and background

Due to the rapid technological development, companies around the world are able to store and analyse huge volumes of information (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2013). Already Eckerson (2002, p. 1) stated that ‘Companies now compete on their ability to absorb and respond to information, not just manufacture and distribute products.’. Stored information is used to support decision-making under uncertainty (Dinter, Lahrmann, & Winter, 2010) and is thus considered to be a competitive advantage. The value of stored information is expressed in the additional benefits due to better decisions (Hilton, 1981; Lawrence, 1999). For example, in customer relationship management (CRM), companies often use stored customer information ‘…to customize offerings and respond to customer needs’ (Mithas, Krishnan, & Fornell, 2005, p. 202), and thus aim at increasing revenues. However, not just information quantity, but also its quality matters.

Information quality is crucial, because poor information quality can result in wrong decisions and economic losses (Gelman, 2010; Heinrich & Klier, 2011; Shankaranarayanan & Cai, 2006). According to a survey by Forbes Insights (2010), information quality problems cost 66% of the companies more than US $2 million per year. Moreover, the assurance of information quality is considered by senior executives to be one of the issues with the highest investment priority in the future (Economist Intelligence Unit, 2011).

Orr (1998) specifies information quality as ‘the measure of the agreement between the data views presented by an information system and that same data in the real world’ (p. 67). A similar definition is presented by Parssian, Sarkar, and Jacob (2004). Information quality is thus recognised as the correspondence between the stored information and its real-world counterpart. Information quality is a multi-dimensional concept and can be expressed, for example, as accuracy, consistency, currency or completeness (Wang & Strong, 1996). In this paper, we focus on currency as one of the most important dimensions (Al-Hakim, 2007; Klein & Callahan, 2007; Lee, Strong, Kahn, & Wang, 2002; Redman, 1996; Wand & Wang, 1996).

A substantial body of literature has been published on measuring and improving information quality and, in particular, currency (Batini, Cappiello, Francalanci, & Maurino, 2009; Heinrich & Hristova, 2014; Heinrich, Kaiser, & Klier, 2007; Heinrich & Klier, 2015; Lee et al., 2002; Parssian et al., 2004; Pipino, Lee, & Wang, 2002; Redman, 1996; Wang, 1998; Woodall, Borek, & Parlikad, 2013). These approaches are generally based on the Total Data Quality Management (TDQM) methodology (Wang, 1998), which consists of four phases: *define, measure, analyse, and improve*. In the first two phases the quality dimension of interest is exactly specified and measured. In the third phase the impact of poor information quality on decision-making is determined. Finally, in the last phase the benefits from improving information quality are compared with the costs of applying quality improvement measures to determine the optimal level of information quality from a net-benefit perspective. In the current paper, we explicitly focus on the second and on the third phases of the TDQM methodology in the context of currency.

Currency can be defined as *the correspondence between the previously correctly stored information and real-world information at the time of measurement* (Heinrich & Klier, 2011, 2015; Pipino et al., 2002; Redman, 1996). Note that in the literature some authors apply this definition to timeliness ('the recorded value is not out of date'; Ballou & Pazer, 1985, p. 153) and some to accuracy ('we observe accuracy and currency as related issues, as we address accuracies that are caused by failures to update data even when changes in the real-world entity require us to do so.'; Wechsler & Even, 2012, p. 1). However, we stick to currency with the above definition for the rest of the paper.

In contrast to accuracy, measuring currency does not require a real-world comparison (Heinrich & Klier, 2011, 2015). Existing metrics for currency (Ballou, Wang, Pazer, & Tayi, 1998; Even & Shankaranarayanan, 2007; Heinrich & Klier, 2011) thus determine ‘an *indication*, not a verified statement’ (Heinrich, Klier, & Kaiser, 2009, p. 5) about the correspondence between the real-world and the previously correctly stored information. To name a few, Heinrich and Klier (2015) interpret currency as the *probability* and Wechsler and Even (2012) as the *likelihood* that the previously correctly stored information is still up to date at the time of measurement. However, decision-makers will benefit more from the application of these metrics if they *in addition* deliver an indication about the current real-world information at the time of measurement, based on the stored information. This can be done by extending the interpretation by Heinrich and Klier (2015) and modelling the temporal change of real-world information.

To illustrate the contribution of this idea, consider a company which possesses a database with the income of its customers stored in the past and which would like to use this information for customer targeting (i.e. customers with higher income would also have higher willingness to pay). In cases where the level of currency measured by the metric of Heinrich and Klier (2015) is very low for some customers, the company may prefer not to use the stored income information at all. However, if the company additionally possesses an indication about the current real-world income at the time of measurement, it may adjust its decision correspondingly and thus increase the value of the stored information.

Therefore, the *first* objective of this paper is to develop an extended metric referring to currency by modelling the temporal change of real-world information, based on the stored information. The contribution of our approach consists in the fact that decision-makers can adjust their decisions, based on the indication about the current real-world information at the time of measurement. This approach is part of the second phase of the TDQM.

As mentioned above, in the third phase of the TDQM methodology the impact of the level of currency on the value of information is determined. A few existing studies quantitatively deal with a similar issue by modelling the dependency between the level of information quality and a utility function (Ballou et al., 1998; Cappiello & Comuzzi, 2009; Even & Shankaranarayanan, 2007; Even, Shankaranarayanan, & Berger, 2007). Among them, Ballou et al. (1998)[1] and Even et al. (2007)[2] assume that lower currency negatively influences the value of information. Even and Shankaranarayanan (2007) argue that quality defects can either reduce the value of information or have no influence on it, and model currency as one of the multipliers that influences information quality. Similarly to Even and Shankaranarayanan (2007), Cappiello and Comuzzi (2009) define an aggregated measure of information quality and consider different utility functions such as the Gaussian function, for example. Most of these approaches define the dependency between the value of information and the level of information quality with a general form. In particular, the authors do not explicitly model the influence of the information quality level on the choice of the decision-maker, although, according to a number of experimental studies (Chengalur-Smith,

---

[1] The authors deal with timeliness.

[2] Here currency is represented by the age of the stored record, determined by the last update.

Ballou, & Pazer, 1999; Fisher, Chengalur-Smith, & Ballou, 2003; Jung, Olfman, L., Ryan, T., & Park, 2005), it may lead to a change in the choice. Note that this choice is different from the choice regarding the application of improvement measures in the fourth phase of the TDQM methodology, which has been extensively discussed in the literature.

Extending the above stream of research to explicitly model the impact of information quality on the choice of the decision-maker is extremely important, because it will provide managers with the missing link between measuring currency and decision-making under uncertainty. Thus, the *second* objective of this paper is to develop such a quantitative tool in the context of currency by extending the normative concept of the value of information (Carter, 1985; Hilton, 1981; Lawrence, 1999; Marschak & Radner, 1972; Repo, 1989). We focus on this concept as it best fits the presented framework (i.e. decision-making under uncertainty).

The paper is structured as follows. In the next section, we present our extended metric referring to currency as the *first* objective of this paper. In section 3, we extend the normative concept of the value of information to explicitly incorporate currency in decision-making under uncertainty as the *second* objective of this paper. In order to demonstrate the feasibility and evaluate the strength of our approach, we apply it to two scenarios from the field of sales management in CRM in section 4. In the last section, we discuss the main implications and limitations of our approach and propose paths for future research.

2. Extended metric referring to currency

As discussed in the Introduction, existing metrics for currency deliver an indication about the correspondence between the stored and real-world information at the time of measurement. This implies that if the metric value is low (i.e. the real-world information may have changed after storage), then the information may be considered worthless for the decision-maker.

To illustrate this idea consider Figure 1, where the change of marital status information over time is exemplified. At the time of storage (three decades after

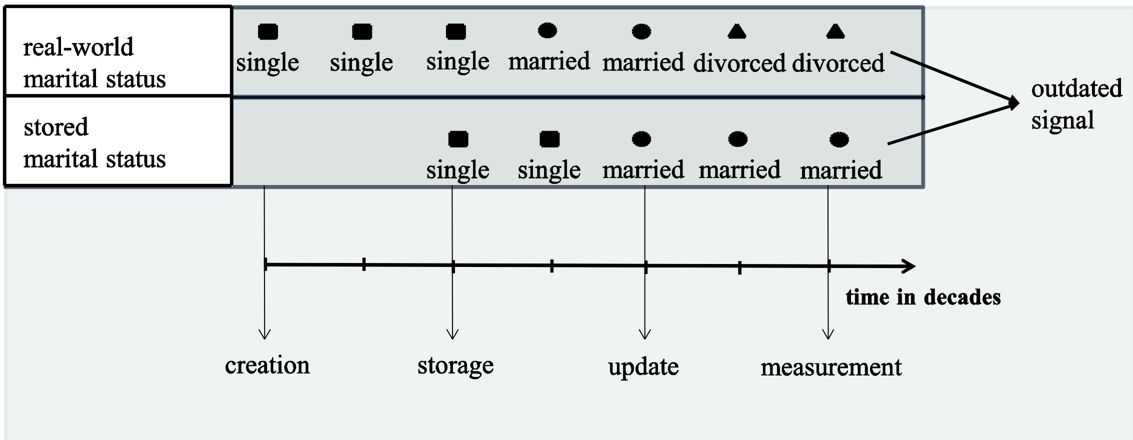

Figure 1. Illustration of the temporal change of real-world information.

creation) the stored and the real-world marital status coincide and both are *single.* Thus, the information is up to date. However, one decade after storage the person gets married and the two values do not coincide anymore. One more decade later the information is updated, so the stored and the real-world marital status coincide again and are both *married*. Then, after one additional decade, the person gets divorced. As a result, at the time of measurement the stored marital status is *married,* while the real-world marital status is *divorced*.

In this case, existing metrics for currency may deliver a low metric value for the correspondence between the stored and the real-world marital status at the time of measurement. Thus, the decision-maker will probably decide not to use the stored information at all to avoid wrong decisions. As a result, the stored information becomes worthless. However, if at the time of measurement, in addition to the low metric value, the decision-maker also knows the distribution of the real-world marital status, based on the stored marital status in the previous periods (e.g. *married* with a probability of 0.2 and *divorced* with a probability of 0.8), then the decision can be adjusted correspondingly. Thus, even though the stored information is outdated, it may still be used to support decisions and is not worthless anymore.

In order to determine the distribution of the real-world information at the time of measurement, we model its temporal change as a stochastic process. A stochastic process $\{Y_t, t \in T\}$ is defined over the probability space $(\Omega, \mathcal{F}, P)$ with the parameter set T, often interpreted as time. For a given $t \in T$, the function $\omega \rightarrow Y_t(\omega), \Omega \rightarrow I$ is a random variable defined over $(\Omega, \mathcal{F}, P)$ with a range I, which can be discrete or continuous. Let $\{Y_t, t \in T\}$ be the stochastic process describing the temporal change of real-world information, where the values[3] $y_t \in I$ of the random variables Y_t are called signals and the range I is called information space. Then Y_0 denotes the random variable describing real-world information (e.g. income, address, marital status) at the time of creation $t = 0$, where it is given by the signal $y_0 \in I$ (e.g. *single*). Real-world information changes over time according to the definition of $\{Y_t, t \in T\}$, and at the time of storage/update $t = t_0 \geq 0$[4] takes the value $Y_{t_0} = y_{t_0}$ (e.g. *married*). Finally, at the time of measurement $t = t_0 + p, p \geq 0$ periods after storage, real-world information is distributed according to the random variable Y_{t_0+p} and the signal y_{t_0+p} is not known to

---

[3] These values differ from the normative concept of the value of information addressed later.

[4] The parameter t_0 denotes the time of storage if no update took place and the time of update otherwise. For simplicity, from now on we will only use storage when referring to it, but will mean both.

the decision-maker. The distribution of Y_{t_0+p} is exactly the indication about the real-world information, which can be used by the decision-maker to support the decision.

In the following, we present two extended metrics to determine the distribution of Y_{t_0+p}. These are the *general form* of the extended metric referring to currency and the *Markov form* of the extended metric referring to currency. If all the real-world signals before the time of measurement are known (i.e. y_0, y_1,…, y_{t_0+p-1}), then the distribution of Y_{t_0+p} can be determined very precisely. This is the *general form* of the extended metric referring to currency. In the above marital status example, this implies that the real-world signals in the periods between creation (i.e. *single*) and storage (i.e. *divorced*) are known.

However, in many realistic situations neither the real-world signals before storage, nor the ones after that are known to the decision-maker. Thus, we also determine the distribution of Y_{t_0+p}, based only on the stored signal y_{t_0} (i.e. *married* in the above example), by making an additional assumption regarding the stochastic process. This is the *Markov form* of the extended metric referring to currency, which is not as precise as the *general form*, but more appropriate for some practical applications. We begin with the *general form* of the extended metric referring to currency.

The *general form* of the extended metric referring to currency is defined for $p \geq 0, t_0 \geq 0$ as:

$$m_g\left(y_{t_0+p} | y_{t_0+p-1}, y_{t_0+p-2}, \dots, y_{t_0}, \dots, y_0, p, t_0\right) :=$$
$$f_{Y_{t_0+p}|Y_{t_0+p-1},Y_{t_0+p-2},\dots,Y_{t_0},\dots,Y_0}\left(y_{t_0+p} | y_{t_0+p-1}, y_{t_0+p-2}, \dots, y_{t_0}, \dots, y_0\right), y_i \in I \;\forall i \geq 0 \quad (1)$$

where $\{Y_t, t \in T\}$ is the general stochastic process describing the temporal change of real-world information, $t_0 \geq 0$ is the time of storage and $p \geq 0$ represents the number of periods between storing the information and measuring currency.

$f_{Y_{t_0+p}|Y_{t_0+p-1},Y_{t_0+p-2},\ldots,Y_{t_0},\ldots,Y_0}\left(y_{t_0+p}\middle|y_{t_0+p-1},y_{t_0+p-2},\ldots y_{t_0},\ldots,y_0\right), y_i \in I\ \forall i \geq 0$ is the conditional probability density (mass) function at the real-world signal y_{t_0+p} for the time of measurement $t_0 + p$, provided that the real-world signals in the periods before the measurement were $y_{t_0+p-1}, y_{t_0+p-2}, \ldots, y_{t_0}, \ldots, y_0$. In addition, $m_g\left(y_{t_0+p}|y_{t_0+p-1}, y_{t_0+p-2}, \ldots, y_{t_0}, \ldots, y_0, p, t_0\right), y_{t_0+p} \in I\ \forall i \geq 0$ is the corresponding metric, which does not result in a single value, but rather a distribution over the possible signals y_{t_0+p} conditioned on the given real-world signals $y_{t_0+p-1}, y_{t_0+p-2}, \ldots, y_{t_0}, \ldots, y_0$.

In the case of discrete information space, the *general form* of the extended metric gives the probabilities for the possible real-world signals at the time of measurement, conditioned on the given real-world signals in the periods before the measurement. In contrast, the metric for currency by Heinrich and Klier (2015) gives the probability that the real-world information did not change between the time of storage t_0 and the time of measurement $t_0 + p$ (i.e. the probability that the stored signal is still up to date). Thus, the contribution of our metric is that it also provides the probabilities for the possible real-world signals at the time of measurement which differ from than the stored one. This is important especially in the case when the correspondence between the stored signal and its real-world counterpart is indicated to be low, because then the decision-maker can adjust his/her decision, based on the probabilities of the *general form* of the extended metric.

The real-world signals $y_{t_0+p-1}, y_{t_0+p-2}, \ldots, y_{t_0}, \ldots, y_0$ before the time of measurement are required as input parameters for the *general form* of the extended metric. However, as mentioned above, in several realistic situations, only y_{t_0} is known to the decision-maker (i.e. neither the history before storage nor the change of information after storage is known). This is similar to the problem described by Heinrich, Klier, and Görz (2012) and Heinrich and Klier (2015), that the time of creation of a signal is often not known. To address this problem, similarly to Heinrich et al. (2012), we consider the time of storage and not the time of creation in determining currency, and restrict the above general stochastic process $\{Y_t, t \in T\}$ to a process which possesses the Markov property (Nelson, 1995; Wechsler & Even, 2012). Thus

$$f_{Y_{t_0+i+1}|Y_{t_0+i},Y_{t_0+i-1},Y_{t_0+i-2},\dots,Y_0}\left(y_{t_0+i+1}\middle|y_{t_0+i}, y_{t_0+i-1}, y_{t_0+i-2}, \dots, y_0\right) =$$

$$f_{Y_{t_0+i+1}|Y_{t_0+i}}\left(y_{t_0+i+1}\middle|y_{t_0+i}\right), \forall i \in [0, p-1] \tag{2}$$

This property implies for $i = 0$ that the change of the real-world information one period after storage depends only on the real-world signal at the time of storage and not on its history before storage. As a result, the signals $y_{t_0-1}, y_{t_0-2}, \dots, y_0$ are not required as inputs anymore. Moreover, by recursively applying Equation (2), we can easily derive the distribution of real-world information at the time of measurement without knowing the real-world signals $y_{t_0+p-1}, y_{t_0+p-2}, \dots, y_{t_0+1}$ in the periods between storage and measurement. If the time of measurement is $p = 2$ periods after storage, then due to Equation (2) the distribution of real-world information Y_{t_0+2} depends on $f_{Y_{t_0+2}|Y_{t_0+1}}$ and the real-world signal at $t_0 + 1$, which is unknown to the decision-maker. However, again due to the Markov property, the distribution of the real-world information Y_{t_0+1} can be derived from $f_{Y_{t_0+1}|Y_{t_0}}$ and the stored signal y_{t_0} which is known to the decision-maker. This recursive idea can analogously be applied for $p > 2$ to estimate $f_{Y_{t_0+p}|Y_{t_0}}\left(y_{t_0+p}\middle|y_{t_0}\right)$.

We define the *Markov form* of the extended metric referring to currency for $p > 0, t_0 \geq 0$ as:

$$m_m\left(y_{t_0+p}|y_{t_0},p,t_0\right) := f_{Y_{t_0+p}|Y_{t_0}}\left(y_{t_0+p}|y_{t_0}\right), y_{t_0+p}, y_{t_0} \in I \tag{3}$$

where $\{Y_t, t \in T\}$ is a stochastic process which possesses the Markov property and $m_m\left(y_{t_0+p}|y_{t_0},p,t_0\right), y_{t_0+p} \in I$ is the corresponding metric, which does not result in a single value, but rather a distribution over the possible signals y_{t_0+p} conditioned on the given stored signal y_{t_0}. The conditional probability density (mass) function of the real-world information Y_{t_0+p} at the time of measurement $t_0 + p$ provided that the stored signal is $Y_{t_0} = y_{t_0}$ (i.e. $f_{Y_{t_0+p}|Y_{t_0}}\left(y_{t_0+p}|y_{t_0}\right)$) is derived recursively as discussed above.

In the following two subsections, we illustrate how the *Markov form* of the extended metric referring to currency can be applied to two scenarios: information with discrete information space such as marital status or address information, and information with continuous information space such as income information. The application of the *general form* of the extended metric referring to currency can be illustrated in a similar fashion.

2.1. Discrete information space scenario

In this subsection, we consider a scenario where the information space consists of a finite number of values. This implies that after storing the signal at t_0 the real-world signal one period after storage can either change to finitely many other values or stay the same. For example, if the marital status of a person is stored as *single*, then one period after storage the real-world marital status can take only one of the values *{single, married, divorced, widowed}*. The same holds for address information, where the number of possible addresses a person could move to one period after storage is high, but still finite. As a result, for any stored signal and each possible value in the information space, the probability that the real-world information takes this value one period after storage can be determined based on publicly available statistical data or other data sources. For example, for the stored *single* marital status, the probabilities that the person got married, divorced, widowed or stayed single one period after storage are determined from the marriage, divorce and mortality rates in a certain country (e.g. from the German Federal Bureau of Statistics in Germany). The same holds for address information, where for each possible address the probability that a person moved to it (or remained there) one period after storage is determined from the residential mobility rates within the same street, district, city, province or country (e.g. also based on data from the German Federal Bureau of Statistics). These probabilities, called transition probabilities, can be determined in a similar fashion for any two consecutive periods after storage.

After their determination, transition probabilities are used to obtain the value of the *Markov form* of the extended metric as follows: let the process $\{Y_t, t \in T\}$ be a Markov chain[5] with an information space $I = \{y^1, \dots, y^n\}, n \geq 1$ consisting of mutually exclusive signals. Then $m_m(y^j|y^k, 1, t_0), j, k \in [1, .., n]$, as defined in Equation (3), is the transition probability from a stored signal y^k at the time of storage t_0 to a real-world signal y^j one period after storage. Similarly, $m_m(y^j|y^s, 1, t_0 + (p-1)), p > 0$ is the transition probability from a stored signal y^s to a real-world signal y^j between any two consecutive periods after storage. To calculate the values of the extended metric $m_m(y^j|y^k, p, t_0)$ $\forall j, \forall k \in \{1, \dots, n\}, t_0 \geq 0, p > 0$, we use the above transition probabilities as inputs to the following recursive model, which is based on the law of total probability:

---

[5] Note that a metric for accuracy, based on Markov chains, is presented in Wechsler and Even (2012).

- $p = 1$: $m_m(y^j|y^k, 1, t_0)$ is determined as described above;
- $p > 1$: $m_m(y^j|y^s, 1, t_0 + (p-1))$ is determined as described above and

  $m_m(y^j|y^k, p, t_0) = \sum_{s=1}^{n} m_m(y^j|y^s, 1, t_0 + (p-1)) m_m(y^s|y^k, p-1, t_0)$

The result from the application of the extended metric is a distribution over the possible real-world signals at the time of measurement, conditioned on the stored signal. This implies that the decision-maker can adjust the decision, if the probability that the stored signal corresponds to the real-world signal is low. For example, if the stored signal is the marital status *single* and it is now indicated that the person is married with high probability at the time of measurement, then the decision should correspond rather to a *married* marital status than to a *single* one (cf. section 4). The same holds for stored address information. If according to the extended metric a person with a given stored address should with high probability have moved to another city, then the new city should be considered in the decision.

We can easily extend the above approach to apply it to a multidimensional information space by using multivariate Markov chains. This is very useful for example in sales campaigns, because often more than one attribute of the customer influences the demand (e.g. gender, education and marital status) and the temporal change of these attributes over time are not independent of each other.

2.2. Continuous information space scenario

In this subsection we consider a scenario where the information space consists of an infinite number of values. This implies that after storing the signal at t_0 the real-world signal one period after storage can either change to infinitely many other values or stay the same. For example, if the stored signal is the income of a person, then the real-world income one decade after storage can take any non-negative value. As opposed to the previous subsection, in this case for each stored signal it is necessary to determine a continuous (transition) probability distribution over the possible real-world signals in the next period. For example, if the stored signal is the income of a given person, then the distribution of the logarithm of the real-world income one period after storage will be normal with expected value depending on the logarithm of the stored income (Guvenen, 2009; Hryshko, 2009). A similar approach is then applied to any two consecutive periods after storage. In order to derive this distribution, historical information is used such as income information of customers or publicly available panel data (SOEP, 2012, cf. section 4).

To determine the value of the *Markov form* of the extended metric, we thus model the process $\{Y_t, t \in T\}$ as an autoregressive process of order one (Brooks, 2008). This implies that for a measurement made one period after storage (i.e. $p = 1$):

$$Y_{t_0+1} = \mu + \phi Y_{t_0} + \varepsilon_{t_0} \tag{4}$$

where $\{\varepsilon_t, t \in T\}$ is a white noise process with variance σ^2 and μ, ϕ are scalars. Thus, the extended metric is the conditional density function of a normally distributed random variable:

$$m_m\left(y_{t_0+1} | y_{t_0}, 1, t_0\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(y_{t_0+1}-(\mu+\phi y_{t_0})\right)^2}{2\sigma^2}} \tag{5}$$

Currency one period after storage can be interpreted as the value of the above density function at y_{t_0} (i.e. $m_m\left(y_{t_0} | y_{t_0}, 1, t_0\right)$). However, as opposed to the previous subsection, since we have a continuous information space here, $m_m\left(y_{t_0} | y_{t_0}, 1, t_0\right)$

cannot be interpreted as probability. In addition, the normal distribution assumption can be easily relaxed by modifying μ, ϕ, and $\{\varepsilon_t, t \in T\}$.

Based on Equation (4), for $p = 2$ the following holds:

$$Y_{t_0+2} = \mu + \phi Y_{t_0+1} + \varepsilon_{t_0+1} = (1+\phi)\mu + \phi^2 Y_{t_0} + \phi\varepsilon_{t_0} + \varepsilon_{t_0+1} \qquad (6)$$

which implies that

$$m_m\left(y_{t_0+2} \mid y_{t_0}, 2, t_0\right) = \frac{1}{\sqrt{2\pi\sigma^2(1+\phi^2)}} e^{-\frac{\left(y_{t_0+2}-\left((1+\phi)\mu+\phi^2 y_{t_0}\right)\right)^2}{2\sigma^2(1+\phi^2)}} \qquad (7)$$

The extended metric for $p > 2$ is analogously estimated:

$$m_m\left(y_{t_0+p} \mid y_{t_0}, p, t_0\right) = \frac{1}{\sqrt{2\pi\sigma^2\frac{1-\phi^{2p}}{1-\phi^2}}} e^{-\frac{\left(y_{t_0+p}-\left(\mu\frac{1-\phi^p}{1-\phi}+\phi^p y_{t_0}\right)\right)^2}{2\sigma^2\left(\frac{1-\phi^{2p}}{1-\phi^2}\right)}} \qquad (8)$$

Therefore, upon measurement the decision-maker knows that the distribution of the real-world signal conditioned on the stored signal is normal with the corresponding expected value and variance and can consider this distribution in the decision. In the example with income, this implies that the distribution of the logarithm of the real-world income at the time of measurement is normal with expected value depending on the logarithm of the stored income, where the parameters of the autoregressive process can be estimated by means of historical data (cf. section 4).

In this section, we presented an extended metric referring to currency and applied it to two scenarios with a discrete and a continuous information space, respectively. In the next section, we present our approach for incorporating this metric in decision-making under uncertainty.

3. Modelling the influence of currency on decision-making under uncertainty

3.1. Development of the model

In this subsection, we model the influence of currency on decision-making under uncertainty by extending the normative concept of the value of information. Thus, we first present this concept. The normative concept of the value of information (VoI) is defined as[6]:

$$VoI := \int_{y\in I} max_{x\in X} \int_{s\in C} w(x,s) f_{S|Y}(s|y) ds f_Y(y) dy - max_{x\in X} \int_{s\in C} w(x,s) f_S(s) ds \qquad (9)$$

where

- S is the random variable describing the states of nature with range (state space) C;
- Y is the random variable describing the signals with range (information space) I;
- $f_{S|Y}$ denotes the conditional density function of the states of nature for a particular signal;
- f_Y represents the density function of the signals;
- f_S stands for the density function of the states of nature;
- X expresses the set of possible choices of the decision-maker;
- $w(x,s)$ is the payoff function for a given choice-state (x,s) combination.

The idea is that the decision-maker has to make an ex ante decision in an uncertain environment without knowing the state of nature that will occur in the future

[6] cf. Hilton(1981) and Lawrence(1999), where up to date information is implicitly assumed.

(i.e. after the instant of the decision). Without additional information, the expected payoff would depend only on the distribution of the states of nature f_S and is exactly the value in the second term of Equation (9; i.e. the subtrahend).

The uncertainty in the environment, which we shall call *environmental uncertainty* from now on, can be reduced by obtaining additional information that indicates which state of nature will occur in the future. This information is represented by the random variable Y, values of which are the signals as discussed in the previous section. If a signal $y \in I$ indicates with certainty which state of nature will occur, it eliminates *environmental uncertainty* completely. However, a signal y can also deliver a 'noisy' indication. This means that it does not indicate one particular state of nature with certainty, but a number of states with their corresponding probabilities (e.g. $\exists s_1, \exists s_2 \in C, s_1 \neq s_2$ with $f_{S|Y}(s_1|y) > 0$ and $f_{S|Y}(s_2|y) > 0$). Then, the signal still reduces *environmental uncertainty*, but does not eliminate it completely. Based on the signal, the decision-maker may change the decision so that it leads to a higher expected payoff. Thus, VoI is calculated as the difference between the expected optimal payoff when additional information is considered and the expected optimal payoff without considering it. Note that for simplicity the definition in Equation (9) assumes that the decision-maker is risk neutral. This is realistic when companies aim only at maximizing their expected profit. However, the model can be easily adjusted for a risk-averse or a risk-seeking decision-maker.

The formula in Equation (9) assumes that the signal corresponds to the real-world one at the time of the decision. However, in many realistic situations this is not the case, as the (stored) signal can be outdated (cf. section 2), and thus the correspondence with the real-world signal is indicated as less than perfect. Then the decision-maker faces another kind of uncertainty in addition to the *environmental* *uncertainty*, which we call *quality uncertainty*. *Quality uncertainty* should be taken into account since it can lead to a change in the ex ante decision and, as a result, to higher ex post value.

We thus extend VoI by explicitly incorporating *quality uncertainty* in the form of our extended metric referring to currency. For illustration purposes, we use the *Markov form* of the extended metric (i.e. Equation 3), but the model can be analogously applied to the *general form* of the metric (i.e. Equation 1).

In Equation (10) we present our model, where VoI after considering *quality uncertainty* ($VoIQ$) depends on the time of storage $t_0 \geq 0$, and on the number of periods p between storage and using the information to support decision-making (i.e. the time of measurement). Moreover, in Equation (11) we define the value of a *signal* (VoS) y_{t_0} stored at time t_0. The reason for this second definition is that a decision is taken based on the observed signal y_{t_0} stored at time t_0 and decisions may differ for different signals resulting in different expected payoffs.

$$VoIQ(t_0, p) \coloneqq$$

$$\int_{y_{t_0} \in I} max_{x \in X} \int_{y_{t_0+p} \in I} \int_{s \in C} w(x,s) f_{S|Y_{t_0+p}}(s|y_{t_0+p}) ds\, m_m(y_{t_0+p}\ |y_{t_0}, p, t_0) dy_{t_0+p} f_{Y_{t_0}}(y_{t_0}) dy_{t_0}$$

$$-max_{x \in X} \int_{s \in C} w(x,s) f_S(s) ds \qquad (10)$$

$$VoS(y_{t_0}, t_0, p) \coloneqq$$

$$max_{x \in X} \int_{y_{t_0+p} \in I} \int_{s \in C} w(x,s) f_{S|Y_{t_0+p}}(s|y_{t_0+p}) ds\, m_m(y_{t_0+p}\ |y_{t_0}, p, t_0) dy_{t_0+p} \qquad (11)$$

The main idea behind Equation (10) is that at $t_0 + p$, the decision-maker observes only the stored signal y_{t_0} and must make a decision based on it (for the *general form* of the extended metric $y_{t_0+p-1}, y_{t_0+p-2}, \dots, y_{t_0+1}, y_{t_0-1}, \dots, y_0$ are additionally taken into account). If the stored signal is indicated to be perfectly up to date (i.e. the corresponding probability in the discrete case is 1), it coincides with the real-world signal and *quality uncertainty* is completely eliminated. However, if the stored signal is indicated not to be perfectly up to date (i.e. the corresponding probability in the discrete case is smaller than 1), then it delivers only a 'noisy' indication about the real-world signal y_{t_0+p} at $t_0 + p$ and *quality uncertainty* exists represented by the *Markov form* of the extended metric $m_m\left(y_{t_0+p} \mid y_{t_0}, p, t_0\right)$. The function $f_{S|Y_{t_0+p}}$ represents *environmental uncertainty* as discussed above.

To illustrate the idea, consider the income example from above. The signal y_{t_0} represents the income of the customer at the time of the storage (e.g. one decade ago). The signal y_{t_0+p} represents the income of the customer at the time of the decision. The function $m_m\left(y_{t_0+p} \mid y_{t_0}, p, t_0\right)$ represents the *quality uncertainty* about the real-world income, and $f_{S|Y_{t_0+p}}$ represents the *environmental uncertainty* regarding the willingness to pay of the customers (i.e. customers with the same income would not necessarily have the same willingness to pay; cf. section 4). Thus, as mentioned above, *quality uncertainty* and *environmental uncertainty* are two different kinds of uncertainty and should both be taken into account by the decision-maker to avoid wrong decisions.

The contribution of our model as opposed to Equation (9) is that for a given stored signal, the choice in the first term of $VoIQ$ may be adapted when considering *quality uncertainty* and will possibly differ from the choice in the first term of VoI (which considers only *environmental uncertainty*), leading to different payoffs. In the next subsection we derive the conditions under which this happens.

3.2. Conditions for a change in the optimal decision when considering currency

As mentioned above, information is valuable because it may result in a different indication about the future states of nature and thus in a change of the decision. The same holds for currency. If after considering it, the indication about future states of nature changes, then so may the decision. In Theorem 1 we provide the conditions under which this happens. The proof of Theorem 1 is provided in the Appendix.

Theorem 1: Let $x^{wo}(y_{t_0}, t_0, p)$ be the optimal choice of the decision-maker at $(t_0 + p)$ for a signal y_{t_0}stored at t_0 without considering currency (cf. Equation 9) and let $x^{w}(y_{t_0}, t_0, p)$ be the optimal choice of the decision-maker at $(t_0 + p)$ for the same stored signal when considering currency (cf. Equation 10). Here we assume that $\frac{\partial^2 w(\mathrm{x,}s)}{\partial^2 x} < 0 \; \forall x \in X, \forall s \in C$ (i.e. the payoff function is concave in the choice of the decision-maker).

Let, in addition,

$$g(s, y_{t_0}, p) := \int_{y_{t_0+p} \in I} f_{S|Y_{t_0+p}}(s|y_{t_0+p}) m_m(y_{t_0+p} \; |y_{t_0}, p, t_0) dy_{t_0+p} \tag{12}$$

$$\mathcal{S}_0(y_{t_0}) := \left\{ s \in C : g(s, y_{t_0}, p) = f_{S|Y_{t_0+p}}(s|y_{t_0}) \right\} \tag{13}$$

Then

(1) $x^{wo}(y_{t_0}, t_0, p) = x^{w}(y_{t_0}, t_0, p)$, if $\mathcal{S}_0(y_{t_0}) = C$

(2) $x^{wo}(y_{t_0}, t_0, p) < x^{w}(y_{t_0}, t_0, p)$, if

$$\int_{s \in C} \frac{\partial w(x^{wo}(y_{t_0}, t_0, p)\,, s)}{\partial x} \left(g(s, y_{t_0}, p) - f_{S|Y_{t_0+p}}(s|y_{t_0}) \right) ds > 0 \tag{14}$$

(3) $x^{wo}(y_{t_0}, t_0, p) > x^{w}(y_{t_0}, t_0, p)$, if

$$\int_{s\in C} \frac{\partial w(x^{wo}(y_{t_0},t_0,p)\,,s)}{\partial x}\Big(g(s, y_{t_0}, p) - f_{S|Y_{t_0+p}}(s|y_{t_0})\Big)\,ds < 0 \qquad (15)$$

To discuss Theorem 1, let us first consider the interpretation of the function $g(s, y_{t_0}, p)$. Since $\int_{s\in C} g(s, y_{t_0}, p)ds = 1$, this function can be interpreted as the distribution of the states of nature, based on the stored signal y_{t_0} after considering both *environmental* and *quality uncertainty.* Thus the set $\mathcal{S}_0(y_{t_0})$ consists of all the states, where the probability of occurrence (for a discrete state space) when considering currency coincides with the probability of occurrence without considering currency. If this is the case for all possible states, then the level of uncertainty (after considering both *environmental* and *quality uncertainty*) stays the same as with only *environmental uncertainty,* and naturally the choice does not change. This is the interpretation behind **(1)**. In contrast, **(2)** and **(3)** state that the solution changes when the probability of occurrence of some states when considering currency changes as opposed to that without considering currency. For example, if a state of nature is much more probable when considering currency and an increase in $x^{w}(y_{t_0}, t_0, p)$ increases the payoff in this state, then $x^{w}(y_{t_0}, t_0, p)$ may be higher than $x^{wo}(y_{t_0}, t_0, p)$.

To illustrate this point, consider again the income example from above. In the case where we consider currency and the real-world income is indicated to be the same as the stored one (and *environmental uncertainty* stays the same), the offer will not change. If the stored income indicates higher real-world income (and *environmental uncertainty* stays the same), then the customer will receive a more expensive offer. If, on the contrary, it indicates lower real-world income (and *environmental uncertainty* stays the same), s/he will receive a cheaper offer. The change in the decision in the last two cases can lead to a change in the relationship between currency and the value of a signal, as discussed in the next subsection.

3.3. The role of the model in the TDQM methodology

In this subsection, we discuss the role of our model in the third phase of the TDQM methodology. In particular, in Figure 2 we illustrate three *possible* dependencies between VoS (i.e. Equation 11) and the level of currency of the stored signal. For illustration purposes, we concentrate in the following on the currency of the stored signal, which represents one of the most important values of the extended metric referring to currency for a discrete information space (cf. section 2). Note, however, that a similar analysis can be done with any other value of the extended metric referring to currency.

Curve 1 implies that VoS is a convex increasing function of currency (i.e. it increases at an increasing rate). This dependency occurs when the decision situation requires a stored signal with a (very) high metric value for a reliable decision (i.e. the indicated correspondence between the stored and the real-world signal must be high). An example for such a case is the stock price for an investor who needs to decide where to invest. There, it is extremely important that the stock price is as up to date as possible.

Similarly, Curve 2 implies that VoS is a concave increasing function of currency and thus increases with currency at a decreasing rate. This dependency occurs when it is important for the decision-maker that the stored information is not too outdated, but it must not necessarily be perfectly up to date for a reliable decision. An example for such a case is the income information for a sales campaign, as discussed above. It is important that the stored income corresponds to some degree to the real-world one, but this correspondence does not have to be perfect for a successful offer.

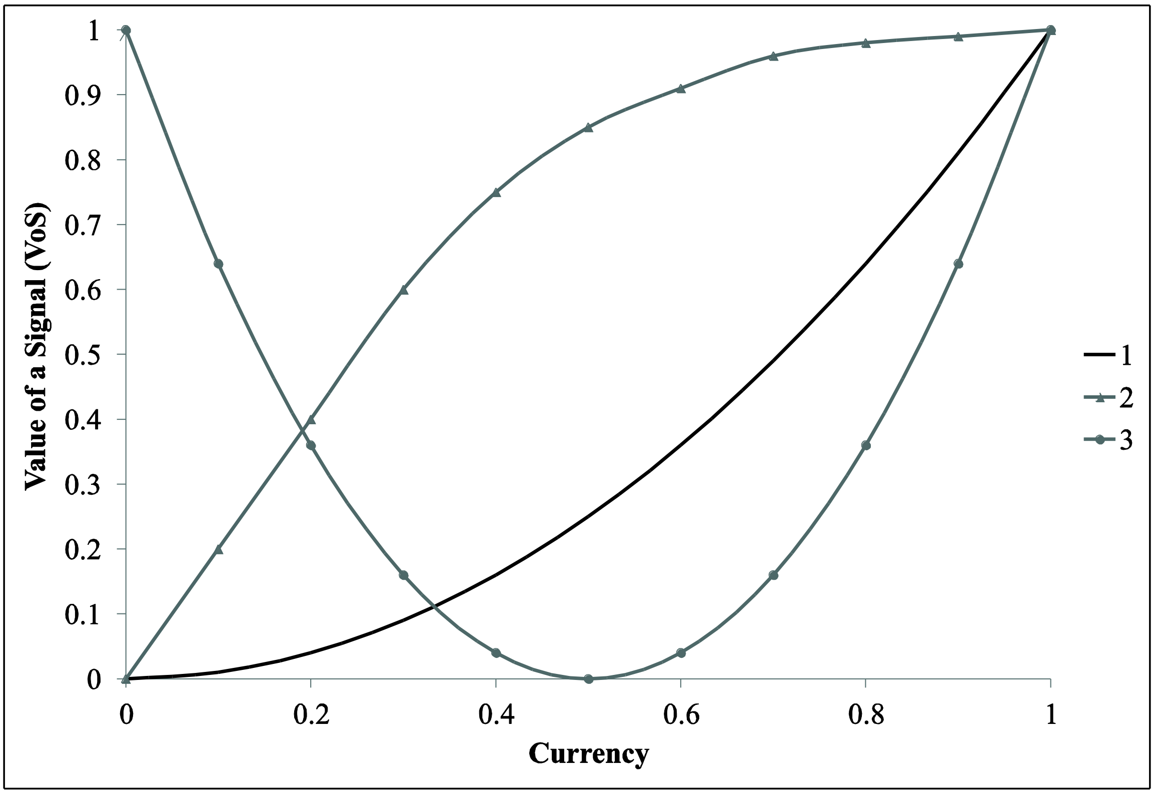

Figure 2. Value of a Signal (VoS) as a function of currency.

In both Curves 1 and 2 VoS is an increasing function of currency, which may support the assumption made by many authors (cf. section 1). As a result, a stored signal with higher currency should also have a higher VoS. However, there is another possible relationship (i.e. Curve 3), which exists in applications but is not considered in the literature. Curve 3 suggests that a stored signal with lower currency may also be valuable, because the decision-maker may learn from the knowledge about the low currency of the stored signal and adjust the decision accordingly. To illustrate this idea, consider a discrete information space consisting of only two signals (i.e. $I = \{y, y'\}$). Let the stored signal be y and the currency of this signal be q (i.e. $m_m(y \mid y, p, t_0) = q$). This implies that $m_m(y' \mid y, p, t_0) = 1 - q$ meaning that the probability that the real-world signal is y' is $(1-q)$. Therefore, if the stored signal is very outdated (i.e. q is very low), then $(1-q)$ is very high and the decision-maker knows almost with certainty that the real-world signal is y', despite the low quality of the stored signal. As a result, s/he can adjust the decision correspondingly, and the value of the stored signal increases. In order to determine this decision, the decision-maker first needs to find the optimal solutions in Equation (10). In the next subsection, we discuss possible approaches for doing this, based on the field of stochastic programming.

3.4. Solution approaches

Our approach for considering currency in decision-making is described by Equation (10), where two optimal choices need to be determined. On the one hand, we need to determine the optimal choice based on the stored signal, by considering both *quality and environmental uncertainty* (i.e. the first term in Equation 10). On the other hand, the optimal choice based *only* on the stored signal and *environmental uncertainty* is required. For both of these cases, in the following, we present possible solution approaches from the field of stochastic programming. Specifically, Equation (10) can be rewritten as:

$$VoIQ(t_0,p) := E_{Y_{t_0}}\left(max_{x\in X} E_{Y_{t_0+p}|Y_{t_0}}\left(E_{S|Y_{t_0+p}}\big(w(x,S)\big)\right)\right) - max_{x\in X} E_S(w(x,S)) \quad (16)$$

where $E_{Y_{t_0+p}|Y_{t_0}}$ is the conditional expectation with respect to the random variable Y_{t_0+p}, modelling the real-world information; $E_{Y_{t_0}}$ is the expectation with respect to Y_{t_0} standing for the stored information; E_S is the expectation with respect to the states of nature; and $E_{S|Y_{t_0+p}}$ is the conditional expectation of the states of nature on the real-world information. The terms $max_{x\in X} E_{Y_{t_0+p}|Y_{t_0}}\left(E_{S|Y_{t_0+p}}\big(w(x,S)\big)\right)$ and

$max_{x \in X} E_S(w(x, S))$ are representative for (unconstrained) stochastic programming problems (Birge & Louveaux, 2011; Gentle, Härdle, & Mori, 2012; Shapiro & Dentcheva, 2014). If these terms can be characterised by a closed-form function of x, then finding a solution would be equivalent to solving an unconstraint (non-)linear optimisation problem, for which there are different approaches in the literature (Hillier & Lieberman, 2004). However, usually this is not the case and other approaches from the field of stochastic programming must be applied.

In the case that no closed-form representation is available and for continuous random variables, the integrals in Equation (16) can be approximated with numerical integration procedures (Birge & Louveaux, 2011) such as quadrature rules for the term $max_{x \in X} E_S(w(x, S))$ and Monte Carlo methods for $max_{x \in X} E_{Y_{t_0+p}|Y_{t_0}}(E_{S|Y_{t_0+p}}(w(x, S)))$ (Robert & Casella, 2010). Another possibility would be to apply stochastic approximation. The basic idea is to stochastically explore X and thus determine the maximum of the objective function. Examples for such approaches are stochastic gradient search and simulated annealing (Gentle et al., 2012; Robert & Casella, 2010). Finally, genetic algorithms, which are based on ideas from natural evolution developments, can also be applied. As opposed to stochastic approximation, genetic algorithms explore X by simultaneously considering a set of possible values instead of only one value (Gentle et al., 2012) and refine this set until the highest fitness is achieved. The presented solution approaches can then be used to determine both optimal choices with and without considering currency and thus $VoIQ(t_0, p)$ in Equation (16). Generally, the selection of the approach depends on the particular mathematical formulation and on the properties of the different functions in Equation (16).This completes the presentation of our approach. In the next section we demonstrate its relevance by applying it to two scenarios from the insurance industry.

4. Evaluation

In this section we evaluate the presented approach, discuss and interpret the results, and explicate relevant practical implications. We concentrate on two scenarios from the field of sales management in CRM for insurance companies. In our analysis we represent customer databases with the panel data from the SOEP (2012), which was made available to us by the German Socio-Economic Panel Study (SOEP) at the German Institute for Economic Research (DIW), Berlin. We chose this application field due to the importance of currency for decision-making there. In particular, when conducting sales campaigns, most insurance companies rely either on their own customer databases or on publicly available information sources and in both cases outdated information may have negative consequences for the success of the campaign.

4.1. General liability insurance scenario

In the first scenario we consider an insurance company, which possesses information regarding the marital status, year of birth (age) and gender of its customers, that was correctly stored in 1998. Thus, the information space is multidimensional, but only the marital status may become outdated. The company is planning to conduct a campaign (in one of the years after 1998) to attract new customers for its general liability insurance by sending different offers per post depending on the marital status of the customers. The company's decision space thus consists of sending one of four offers (i.e. for single, married, divorced or widowed customers) or not sending an offer at all. The customer can either accept or reject the corresponding offer. An offer can be rejected because the customer would not like any general liability insurance. The payoff function is the net profit per person and is calculated based on the annual reports of a number of major insurance companies in Germany as well as on the corresponding postal and labour costs.

In order to demonstrate the relevance of our metric, for a given age and gender of the customers, we first measure the currency of the marital status information. Currency is estimated following the interpretation by Heinrich and Klier (2015) with publicly available statistical information about the marriage, divorce and mortality rates in Germany for the years 1998-2009 (German Federal Bureau of Statistics). In Figure 3, the results for the signal ‘divorced female’ are presented. For better illustration, we plotted only the currency for selected groups of customers (e.g. age between 18 and 30). As we can see, the currency of the marital status information decreases over time and it becomes less than 50% in 2009. This implies that in 2009 the probability that the stored signal is still up to date is lower than the probability that it is not. Therefore, the company may choose not to consider the stored marital status signal at all. However, if the company possesses an indication about the probability that the customer is married or widowed at the time of the campaign, it can adjust the decision and send the corresponding offer.

To determine this indication we apply the *Markov form* of the extended metric referring to currency presented in section 2, and in particular the approach in subsection 2.1 because the information space is discrete. Note that a customer who is married, divorced or widowed can never become single again, and a single customer cannot become divorced or widowed one period later. Table 1 presents the results for a female customer, aged between 18 and 30, whose marital status was stored as ‘divorced’ in 1998. The year of the decision is 2009. As we can see, the probability that such a customer got married in the meantime is in all cases higher than the one that she stayed divorced, which should be taken into account by the decision-maker. As the example shows, the extended metric gives the probabilities for all possible real-world signals at the time of measurement. This demonstrates the *first* objective of our approach.

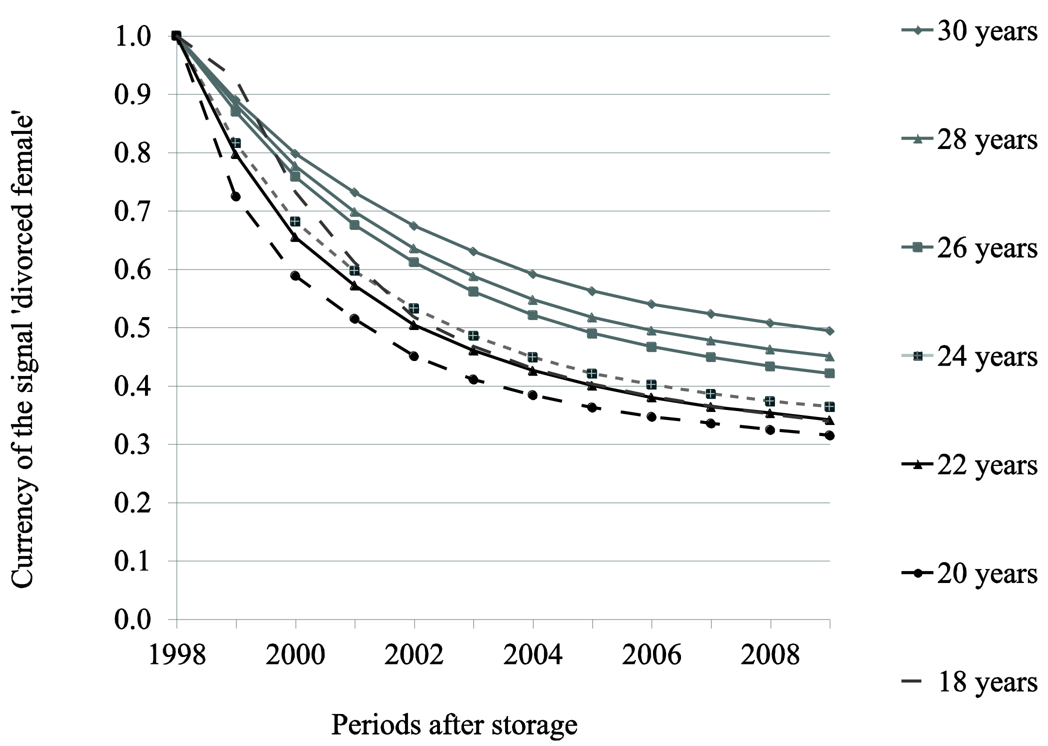

Figure 3. Currency of the signal ‘divorced female’.

The results from the extended metric can now be considered in the decision of the company according to the approach presented in section 3. We thus apply it to the panel data from SOEP (2012) for the years 1998 and 2009, where for each person his/her gender, age and marital status (for each of the 2 years) are given. Based on the stored ‘single’ and ‘divorced’ signals for people aged 18 to 30, we calculate the optimal decisions in 2009 with and without considering currency (i.e. the two ex ante decisions per person), and the optimal decision for the current real-world signal in 2009 (i.e. the ex post decision). Finally, we calculate the ex post loss per person by comparing the payoffs resulting from the ex ante and the ex post decisions.

The results are as follows: changing the ex ante decision based on the knowledge about the currency of the stored information leads to significantly lower ex post loss per person (69% less) for people whose marital status was single in 1998. From the 980 people who were single in 1998, 522 were not single anymore in 2009,

Table 1. Probability distribution for the stored signal ‘divorced female’ customer (2009).

Marital status/Age	**30 years**	**28 years**	**26 years**	**24 years**	**22 years**	**20 years**	**18 years**
Married	≈0.51	≈0.55	≈0.58	≈0.64	≈0.66	≈0.68	≈0.66
Divorced	≈0.49	≈0.45	≈0.42	≈0.36	≈0.34	≈0.32	≈0.34
Widowed	≈0	≈0	≈0	≈0	≈0	≈0	≈0

implying that, based on the ex ante decision *without considering currency*, they received a wrong offer. From them, 488 were married with a loss of 129€ per person; 32 were divorced with a loss of 149.1€ per person; and two were widowed with a loss of 112.9€ per person. The wrong ex ante decisions *when considering currency* were as follows: From the (980 – 522 =) 458 singles in 1998 who were still single in 2009, 140 received a wrong offer for married ones resulting in a loss of 92.2€ per person. From the 488 singles in 1998 who got married by 2009, 292 received an offer for singles (a loss of 129€ per person). From the 32 singles in 1998 who got divorced by 2009, 14 received an offer for married ones (a loss of 108.3€ per person) and 18 received an offer for singles (a loss of 149.1€ per person). Finally, both of the singles in 1998 who were widowed by 2009 received an offer for married people (a loss of 144.5€ per person).

This results in a total loss of approx. 150,053€ without considering currency and a total loss of approx. 88,953€ when considering currency.

The same holds for young people who were divorced in 1998. They should receive an offer for married customers, because for them the probability that they are married in 2009 is higher than the probability that they are still divorced (cf. Table 1). If currency is not considered, they will receive an offer for divorced customers. Again, considering currency leads to significantly lower ex post loss per person (91% less), as opposed to not considering it. As the example shows, currency can be adequately considered in decision-making based on the extended model in section 3, and not doing so leads to wrong decisions and ex post losses. This demonstrates the *second* objective of our approach.

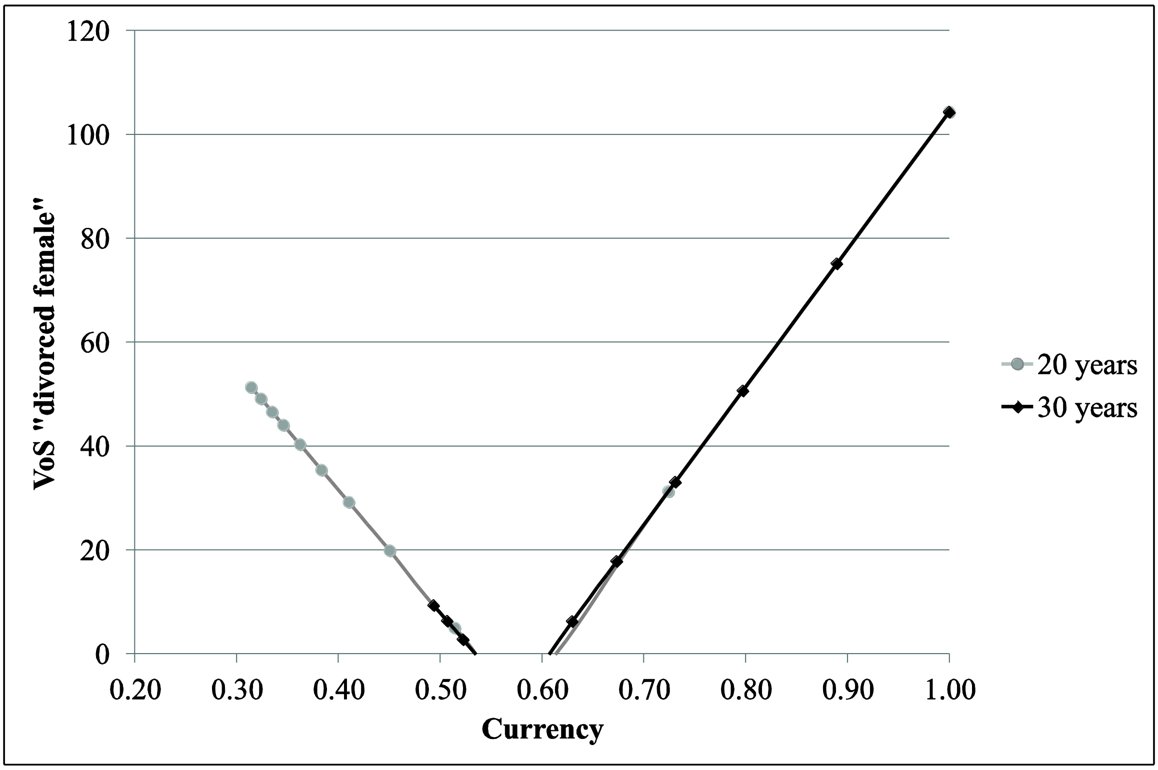

Figure 4. Value of a Signal (VoS) of ‘divorced female’ in different age groups.

In order to show the role of our approach for the third phase of the TDQM methodology, we calculate the VoS of ‘divorced female’ for the years 1998 to 2009 and examine the dependency between this value and the level of currency. The results for the females aged 20 and 30 are shown in Figure 4. Note that the dependency is similar in the other age and gender groups. Naturally, VoS decreases with decreasing currency (cf. Figure 3). However, as soon as the optimal decision changes from sending an offer for divorced to sending an offer for married customers (this happens at a currency of approx. 0.6), VoS increases (cf. Curve 3 in the subsection 3.3), although currency continues to decrease. This is because the decision-maker learns from the knowledge about the low currency of the stored signal and adjusts the decision accordingly. As a result, even information with low currency can be valuable. This effect cannot be explained with existing approaches as they do not consider the change in the choice of the decision-maker when modelling the influence of the level of currency on VoS.

4.2. Pension insurance scenario

In this subsection, we consider a scenario where an insurance company would like to attract new customers for its pension insurance. It again possesses customer information and would like to make an individual offer to each customer who does not have the insurance yet. The aim of the company is to offer a premium that is the closest to the premium the customer is willing to pay for the insurance. Therefore, the insurance company plans to use the annual income of the customer as additional information for specifying the best offer. However, this information is not necessarily up to date as it was stored some time ago. In the following, we describe the scenario in detail.

To begin with, the payoff function is given by $w(x, s) = -c(x - s)^2, c > 0$, where c represents the cost for making the wrong offer, x stands for the premium of the corresponding offer and s is the premium the customer is willing to pay. This implies that by maximising the expected profit the company aims at *minimising* the expected loss. This kind of loss function is often used in the literature to model the production decision of a supplier (Gavirneni, Kapuscinski, & Tayur, 1999). If on the one hand, a supplier produces less than the actual demand, s/he usually has additional costs either due to contract agreement or as reputational costs. If on the other hand, a supplier produces more than the actual demand s/he again has to bear additional costs, either due to storage costs or for production costs of non-storable goods. We suppose for simplicity that the costs for over- and underestimating the customer's willingness to pay are the same.

The pension insurance is especially profitable for customers with high income. The idea is that wealthy customers would like to maintain their standard of living even after they retire and would therefore pay a higher premium to receive a higher pension (Klier, Heidemann, & Günther, 2010). However, since customers may choose to pay different amounts of money, even if they have the same income, there is *environmental uncertainty* regarding the exact premium paid.

To model this *environmental uncertainty* we use a skew normal distribution, which allows us to represent high willingness to pay for wealthy customers, medium willingness to pay for the average earners and low willingness to pay for low earners. In Figure 5 we plotted the functions for three types of customers, depending on their annual income, where a customer earning approx. 42,000€ is considered to be an average earner in Germany for 2011 (SOEP, 2012). The parameters of the skew normal distribution were estimated based on the idea that in Germany up to 4% of the annual income for additional pension insurance is subsidised. In Figure 5 a customer earning 30,000€ per year is naturally willing to pay less with higher probability as compared to a customer earning 50,000€ per year.

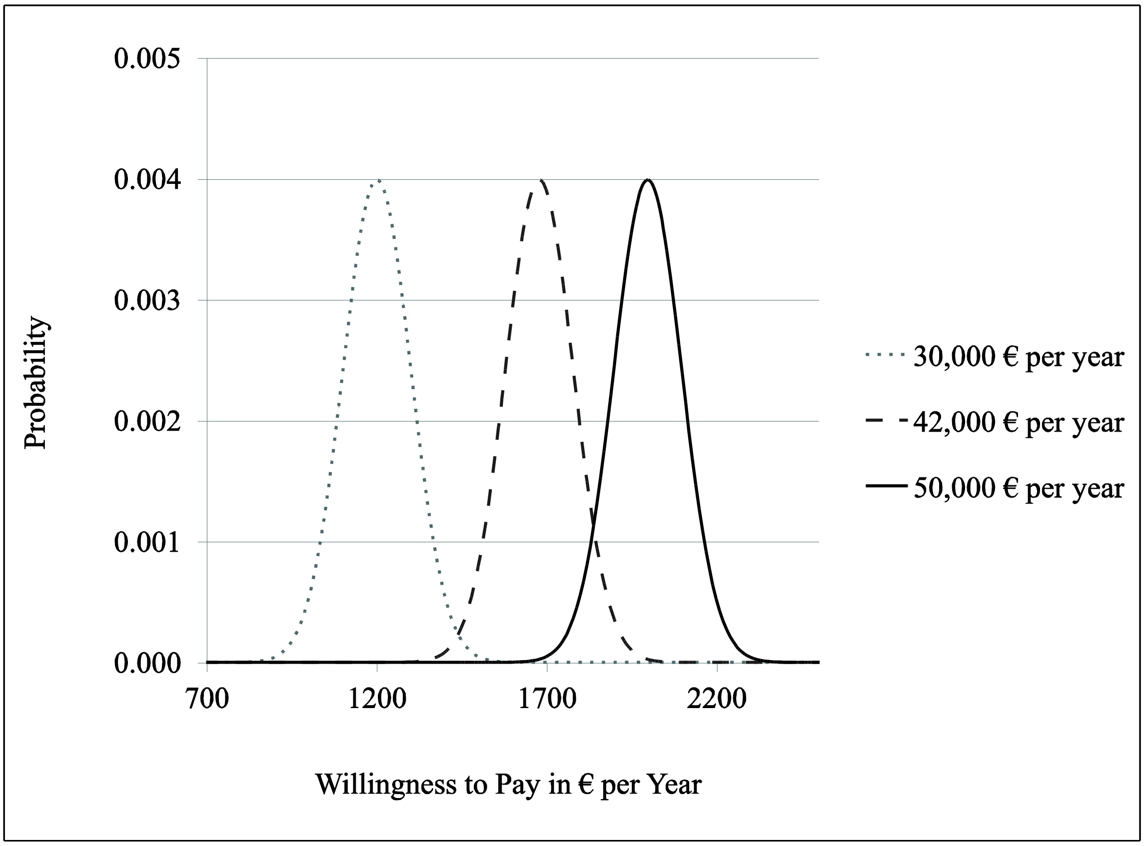

Figure 5. *Environmental uncertainty* according to annual income.

As mentioned above, the stored income may not necessarily be up to date, and so *quality uncertainty* exists in addition to *environmental uncertainty*. As opposed to the previous subsection, here the information space is continuous and therefore the approach in subsection 2.2 is applied to consider currency. In order to estimate the parameters of the autoregressive process, we used the logarithm of historical income data (Guvenen, 2009; Hryshko, 2009) from the SOEP (2012) for the years 1991 till 2011. This was done based on a randomly extracted data set, which was excluded from further evaluation.

Figure 6 plots the conditional distribution of the logarithm of the real-world income for different periods p after storage, given that the stored log-income was the average for the year 1991 (i.e. approx. 9.7, which is equivalent to approx. 16,570€ per year). If we compare the distributions of the logarithm of the real-world income $p = 1$

and $p = 10$ periods after storage, we can clearly see that, while the distribution in the second case (i.e. low currency) is relatively flat and indicates many possible values (including the stored log-value of approx. 9.7), the one in the first case (i.e. high currency) indicates with a relatively high probability that the real-world log-income equals the stored one. This illustrates that lower currency increases the uncertainty between the stored log-income information and the real-world information. Moreover, since this holds for the distribution of the log-income, the distribution of the income itself will be generally characterised by even higher uncertainty. If *currency* is not taken into account, the company may make a premium offer that strongly deviates from the true income of the customer, and will incur the corresponding costs (cf. Figure 7 below). This demonstrates the *first* objective of our approach.

To demonstrate the second objective of our approach, we applied it to the income information for 1991 and 2011 of the individuals in SOEP (2012). We compared the ex ante decisions in 2011, based on the stored signal in 1991, with and without considering currency with the ex post decision, based on the real-world signal in 2011. Note that we truncated the distribution of the real-world log-income for the highest value in 2011 and zero. In Table 2, the main descriptive statistics of the difference between the two optimal decisions are provided. In the last row, the p-value for the two sample t-test is given. The results indicate that by considering currency, the insurance company would significantly adjust the premium offer as opposed to not doing so.

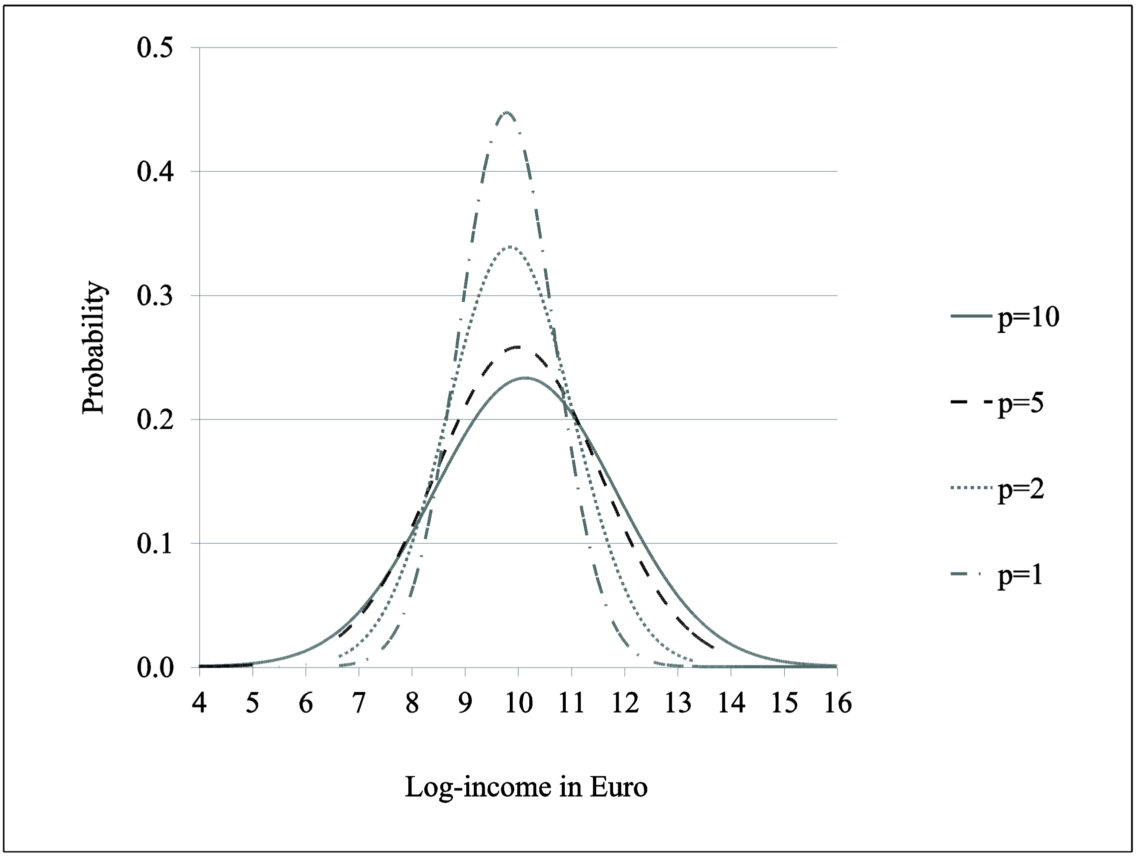

Figure 6. Distribution of the real-world log-income.

Finally we calculated the ex post loss for each of the two ex ante decisions as compared to the ex post decision for 2011. The results are presented in Figure 7. For a better illustration we classified them according to the value of the stored income. For example, the category $\in [q20, q40)$ consists of all the cases in which the stored income was between the 20% quantile and the 40% quantile of the data. We can see that when considering currency the insurance company would adjust its decision, and thus always incur a lower ex post loss per person as opposed to not doing so. This demonstrates the *second* objective of our approach.

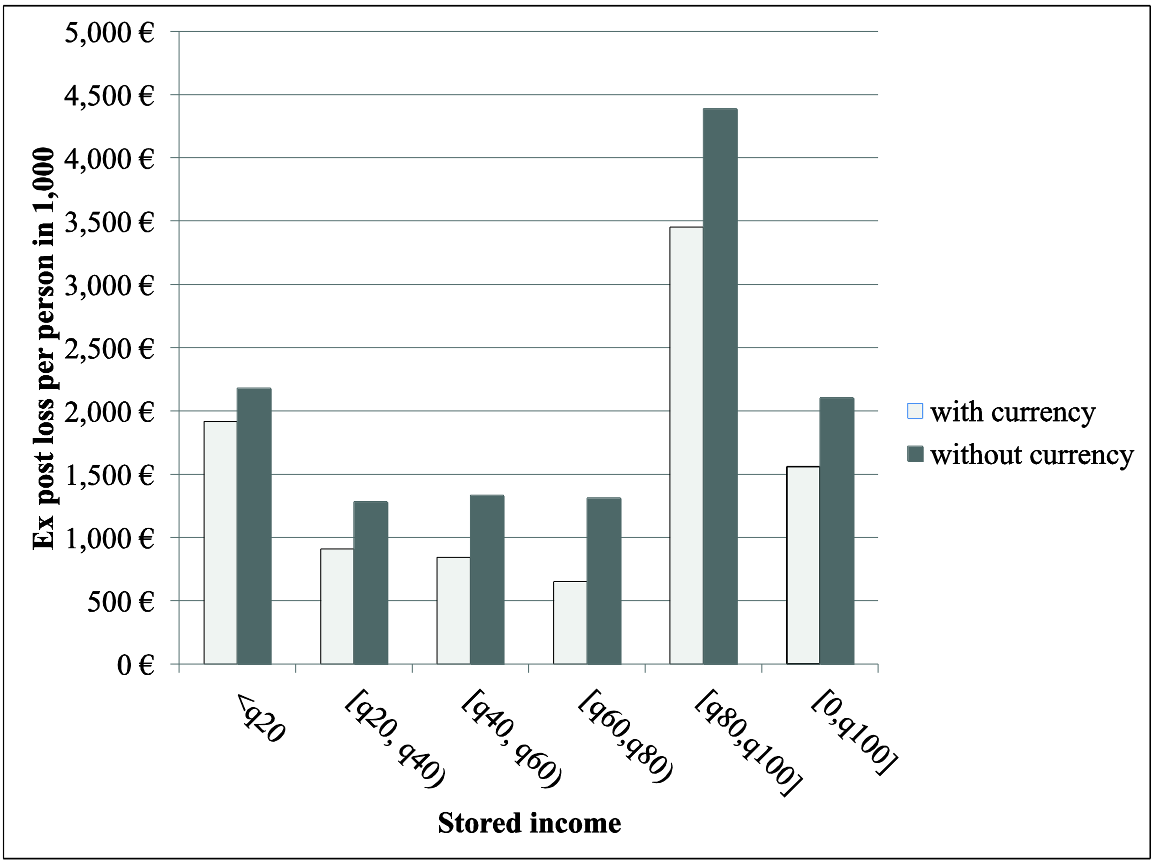

Figure 7. Ex post loss per person with and without considering currency.

Table 2. Descriptive statistics of the difference between the two optimal decisions.

Statistics	Value
Minimum	33€
Maximum	2344€
Mean	877€
Median	≈ 870€
Variance	127,246€
p-value of the two-sample t-test (H_0: $mean_1 = mean_2$)	p-value ≈ 0

4.3.Practical implications

In this section we discuss some practical implications stemming from the two empirical evaluations:

- *Quality uncertainty is relevant in real-world decision-making.* Both the marital status information and the income information of customers can often be outdated. As we demonstrated, based on our extended metric referring to currency, *quality uncertainty* can be measured and modelled. This requires historical information, which can be obtained either from publicly available sources (e.g. the German Federal Bureau of Statistics) or from private sources such as customer databases. Other possibilities are presented in Heinrich et al. (2012).
- *Currency influences the ex ante choice of the decision-maker significantly.* We demonstrated that in both scenarios the choices with and without considering currency may significantly differ from each other. The reason for this is that considering currency results in a different indication for the states of nature than not doing so (cf. subsection 3.2). Thus, companies should take currency into account to avoid wrong decisions from an ex ante point of view.
- *Incorporating quality uncertainty leads to more valuable ex post decisions than not doing so.* This was the case in both scenarios. Thus, companies should consider *quality uncertainty* in their decisions to avoid or reduce ex post losses.
- *The choice of the decision-maker should be considered in the third phase of the TDQM methodology.* If this is not done, the value of information can be underestimated. This can lead to wrong decisions in the next, fourth phase of the

TDQM methodology, such as investing in quality improvement measures, where the benefits in the form of an increase in the value of information do not outweigh the costs for the measure.

4.4. Further application domains

In the previous subsections we demonstrated the application of our approach to two business-to-consumer (B2C) scenarios in the area of sales management. Generally, this approach is not restricted to this field and can be applied to further scenarios in which the necessary model inputs can be determined. Such inputs are the decision space, the states of nature and their distribution, the payoff function, the signals and their distribution (both conditional and unconditional), and the historical data regarding the temporal change of information. To demonstrate the broader scope of our approach, in the following we consider a business-to-business (B2B) CRM scenario in the fields of product co-development and customer retention (Winer, 2001) and describe how currency can be measured with the extended metric and considered in decision-making. A B2B scenario would usually concern fewer products and be longer lasting and more individualised than a B2C one (Reed, Story, & Saker, 2004).

To illustrate the idea, consider an electronic component supplier (e.g. of sensors, semiconductors, cf. Homburg, Wilczek, & Hahn, 2014) that has developed a new product and would like to offer a free trial to some of its important customers by additionally asking them for their feedback regarding this product. The aim is, on the one hand, to increase the loyalty of these particular customers, and, on the other hand, to improve the product by considering customers' feedback. The decision is thus whether to contact a particular customer or not. To make the best possible choice, the company uses historical data on the purchasing history of the customers and determines the payoff function, based on the customer lifetime value (Gupta et al., 2006). In particular, the customers with the highest customer lifetime value also have the highest net payoff and are considered to provide the most valuable feedback. Thus, the states of nature are described by the customer retention rates and purchasing behaviour. Both of these depend on many factors such as company size, variety of products, customer base, etc. which represent the signal space and are known (e.g. stored in a customer database). However, some of these signals may be outdated and may thus lead to a wrong customer lifetime value, and, as a result, wrong decisions. To avoid such cases, for each customer the currency of the significant signals can to be modelled with the extended metric for currency. To initialise the metric, publicly available data on other companies with the same customer profile may be applied. This completes the evaluation of our approach. In the next section, the main conclusions are drawn and paths for future research are discussed.

5. Conclusion, limitations and future research

In this paper, we present an extended metric referring to currency and a quantitative approach to model the influence of currency on decision-making under uncertainty. Our approach is based on the normative concept of the value of information. We extended both the research stream which deals with measuring currency as a second phase of the TDQM methodology, and the research stream which focuses on modelling the influence of currency on the value of information as a third phase of the TDQM methodology. We addressed the objectives listed in the Introduction as follows:

The *first* objective in this paper was to develop an extended metric referring to currency, which provides not only an indication about the correspondence between the stored and the real-world information, but also an indication about the real-world information at the time of measurement. We fulfilled this objective by proposing both a *general form* metric and a *Markov form* metric, based on the Markov property of stochastic processes. Moreover, we presented two examples for discrete and continuous state spaces, which were applied to real-world scenarios in section 4.

The *second* objective of this paper was to provide decision-makers with a quantitative tool to incorporate the level of currency in decision-making. This objective was fulfilled by extending the normative concept of the value of information. Our model allows the calculation of the value of information by considering that the choice of the decision-maker may change if s/he knows the level of currency of the stored information, and we derived the conditions under which this happens. Moreover, we demonstrated that considering this choice may result in a dependency between the value of information and currency which is not modelled by existing approach. Finally, we empirically demonstrated that our approach leads to better ex post decisions than not considering currency.

The main limitation of our approach is that determining the extended metric referring to currency usually requires a substantial amount of data for a reliable estimation, which is not always given in reality. We demonstrated one possible way to do this, based on historical data. A list of other possible methods is given in Heinrich et al. (2012).

Another limitation is that our approach is restricted to the information quality dimension currency. Since low levels of other information quality dimensions such as consistency may just as well result in wrong decisions and economic losses, future research should aim at extending our approach to model the influence of other dimensions on decision-making and on the value of information.

Finally, existing information quality metrics measure only one information quality dimension. This implies that with the current state of the art, our model can only consider one dimension (i.e. currency) at a time. However, decision-makers often possess information about the level of quality according to different dimensions and assign different levels of importance to them. Future works should therefore focus on developing our approach further to account for the simultaneous effect of multiple information quality dimensions on decision-making and on the value of information.

References

Al-Hakim, L. (2007). Information quality factors affecting innovation process. *International Journal of Information Quality, 1*(2), 162–176.

Ballou, D., Wang, R., Pazer, H., & Tayi, G. K. (1998). Modeling information manufacturing systems to determine information product quality. *Management Science, 44*(4), 462–484.

Ballou, D. P., & Pazer, H. L. (1985). Modeling data and process quality in multi-input, multi-output information systems. *Management Science, 31*(2), 150–162.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys (CSUR), 41*(3), 16.

Birge, J. R., & Louveaux, F. (2011). *Introduction to stochastic programming*: Springer Science & Business Media.

Brooks, C. (2008). *Introductory econometrics for finance*. Cambridge: Cambridge University Press.

Cappiello, C., & Comuzzi, M. (2009). A utility-based model to define the optimal data quality level in IT service offerings. In *Proceedings of the 17th European*

Conference on Information Systems (ECIS) (Paper 76). Verona, Italy.

Carter, M. P. (1985). The valuing of management information Part I: The Bayesian approach. *Journal of Information Science, 10*(1), 1–9.

Chengalur-Smith, I. N., Ballou, D. P., & Pazer, H. L. (1999). The impact of data quality information on decision making: an exploratory analysis. *IEEE Transactions on Knowledge and Data Engineering, 11*(6), 853–864.

Dinter, B., Lahrmann, G., & Winter, R. (2010). Information Logistics as a Conceptual Foundation for Enterprise-Wide Decision Support. *Journal of Decision Systems, 19*(2), 175–200.

Eckerson, W. (2002) *Data Warehousing Special Report: Data quality and the bottom line*. Applications Development Trends May 2002.

Economist Intelligence Unit. (2011). *Big data: Harnessing a game-changing asset*. Retrieved from http://www.sas.com/resources/asset/SAS_BigData_final.pdf

Even, A., & Shankaranarayanan, G. (2007). Utility-driven assessment of data quality. *ACM SIGMIS Database, 38*(2), 75–93.

Even, A., Shankaranarayanan, G., & Berger, P. D. (2007). Economics-driven data management: An application to the design of tabular data sets. *IEEE Transactions on Knowledge and Data Engineering, 19*(6), 818–831.

Fisher, C. W., Chengalur-Smith, I., & Ballou, D. P. (2003). The impact of experience and time on the use of data quality information in decision making. *Information Systems Research, 14*(2), 170–188.

Forbes Insights. (2010). *Managing Information in the Enterprise: Perspectives for Business Leader*. Retrieved from http://images.forbes.com/forbesinsights/StudyPDFs/SAP_InformationManagement_

04_2010.pdf

Gavirneni, S., Kapuscinski, R., & Tayur, S. (1999). Value of information in capacitated supply chains. *Management Science, 45*(1), 16–24.

Gelman, I. A. (2010). Setting priorities for data accuracy improvements in satisficing decision-making scenarios: A guiding theory. *Decision Support Systems, 48*(4), 507–520.

Gentle, J. E., Härdle, W. K., & Mori, Y. (2012). *Handbook of computational statistics: concepts and methods*: Springer Science & Business Media, Berlin Heidelberg.

Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Sriram, S. (2006). Modeling customer lifetime value. *Journal of Service Research, 9*(2), 139–155.

Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic Dynamics, 12*(1), 58–79.

Heinrich, B., & Hristova, D. (2014). A Fuzzy Metric for Currency in the Context of Big Data. In *Proceedings of the 22nd European Conference on Information Systems (ECIS)* . Tel Aviv, Israel.

Heinrich, B., Kaiser, M., & Klier, M. (2007). How to measure Data Quality? A Metric-based Approach. In *Proceedings of the 28th International Conference of Information Systems (ICIS)* .

Heinrich, B., & Klier, M. (2011). Assessing data currency—a probabilistic approach. *Journal of Information Science, 37*(1), 86–100.

Heinrich, B., & Klier, M. (2015). Metric-based Data Quality Assessment - Developing and Evaluating a Probability-based Currency Metric. *Decision Support Systems, 72,* 82–96.

Heinrich, B., Klier, M., & Görz, Q. (2012). A metric-based approach for measuring the currency of data in information systems (in German). *Journal of Business Economics, 82*(11), 1193–1228.

Heinrich, B., Klier, M., & Kaiser, M. (2009). A procedure to develop metrics for currency and its application in CRM. *Journal of Data and Information Quality (JDIQ), 1*, 1–28.

Hillier, F. S., & Lieberman, G. J. (2004). *Introduction to Operations Research* (8th ed.): Mcgraw-Hill Publ.Comp, New York.

Hilton, R. W. (1981). The determinants of information value: Synthesizing some general results. *Management Science, 27*(1), 57–64.

Homburg, C., Wilczek, H., & Hahn, A. (2014). Looking Beyond the Horizon: How to Approach the Customers' Customers in Business-to-Business Markets. *Journal of Marketing, 78*(5), 58–77.

Hryshko, D. (2009). *RIP to HIP: The data reject heterogeneous labor income profiles*. University of Alberta, Edmonton.

Jung, W., Olfman, L., Ryan, T., & Park, Y. (2005). An experimental study of the effects of contextual data quality and task complexity on decision performance. In *Proceedings of the 2005 IEEE International Conference on Information Reuse and Integration (IRI 2005)* (pp. 149–154). Las Vegas, Nevada.

Klein, B. D., & Callahan, T. J. (2007). A comparison of information technology professionals' and data consumers' perceptions of the importance of the dimensions of information quality. *International Journal of Information Quality, 1*(4), 392–411.

Klier, M., Heidemann, J., & Günther, B. (2010). The determination of the customer potential in controlling (in German). *Controlling & Management Review, 54*(1), 48–54.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2013). Big data, analytics and the path from insights to value. *MIT Sloan Management Review, 21*.

Lawrence, D. B. (1999). *The economic value of information*. New York: Springer.

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. *Information & management, 40*(2), 133–146.

Marschak, J., & Radner, R. (1972). *Economic theory of teams*. New Haven and London: Yale University Press.

Mithas, S., Krishnan, M. S., & Fornell, C. (2005). Why do customer relationship management applications affect customer satisfaction? *Journal of Marketing, 69*(4), 201–209.

Nelson, R. (1995). *Probability, stochastic processes, and queueing theory: the mathematics of computer performance modeling*. New York, NY: Springer.

Orr, K. (1998). Data quality and systems theory. *Communications of the ACM, 41*(2), 66–71.

Parssian, A., Sarkar, S., & Jacob, V. S. (2004). Assessing data quality for information products: impact of selection, projection, and Cartesian product. *Management Science, 50*(7), 967–982.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM, 45*(4), 211–218.

Redman, T. C. (1996). *Data Quality for the Information Age*. Boston, MA: Artech House.

Reed, G., Story, V., & Saker, J. (2004). Business-to-business marketing: What is important to the practitioner? *Marketing Intelligence & Planning, 22*(5), 501–510.

Repo, A. J. (1989). The value of information: Approaches in economics, accounting, and management science. *Journal of the American Society for Information Science, 40*(2), 68–85.

Robert, C., & Casella, G. (2010). *Introducing Monte Carlo Methods with R*: New York, NY: Springer Science+ Business Media, LLC.

Shankaranarayanan, G., & Cai, Y. (2006). Supporting data quality management in decision-making. *Decision Support Systems, 42*(1), 302–317.

Shapiro, A., & Dentcheva, D. (2014). *Lectures on stochastic programming: modeling and theory* (Vol. 16): SIAM, Philadelphia.

SOEP (2012). Socio-economic panel stody (SOEP), Data for years 1984-2011, Version 28, SOEP, 2012.

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM, 39*(11), 86–95.

Wang, R. Y. (1998). A product perspective on total data quality management. *Communications of the ACM, 41*(2), 58–65.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5–33.

Wechsler, A., & Even, A. (2012). Using a Markov-Chain model for assessing accuracy degradation and developing data maintenance policies. In *Proceedings of 18th Americas Conference on Information Systems* . Seattle, WA.

Winer, R. S. (2001). A framework for customer relationship management. *California management review, 43*(4), 89–105.

Woodall, P., Borek, A., & Parlikad, A. K. (2013). Data quality assessment: The Hybrid Approach. *Information & management, 50*(7), 369–382.

Appendix: Proof of Theorem 1

Proof of Theorem 1: Due to the concavity assumption, the two optimal solutions need only to satisfy the first-order conditions:

$$\int_{s\in C} \frac{\partial w(x^{wo}(y_{t_0},t_0,p),s)}{\partial x} f_{S|Y_{t_0+p}}\left(s\middle|y_{t_0}\right)ds = 0 \tag{17}$$

$$\int_{y_{t_0+p}\in I}\int_{s\in C} \frac{\partial w(x^{w}(y_{t_0},t_0,p)\,,s)}{\partial x} f_{S|Y_{t_0+p}}\left(s\middle|y_{t_0+p}\right)ds\, m_m\left(y_{t_0+p}\,|y_{t_0},p,t_0\right)dy_{t_0+p} = 0 \tag{18}$$

Note that Equation (18) is equivalent to:

$$\int_{s\in C} \frac{\partial w(x^{w}(y_{t_0},t_0,p)\,,s)}{\partial x} g(s,y_{t_0},p)ds = 0 \tag{19}$$

This implies that $x^{wo}\left(y_{t_0},t_0,p\right) = x^{w}(y_{t_0},t_0,p)$ will hold, if $x^{wo}\left(y_{t_0},t_0,p\right)$ satisfies Equation (19).

(1) Let $\mathcal{S}_0(y_{t_0}) = C$ implying that $g\left(s,y_{t_0},p\right) = f_{S|Y_{t_0+p}}\left(s\middle|y_{t_0}\right)\ \forall s \in C$. Then Equation (19) is equivalent to Equation (17) and thus $x^{wo}\left(y_{t_0},t_0,p\right) = x^{w}(y_{t_0},t_0,p)$.

(2) Since

$$\int_{s\in S} \frac{\partial w(x^{wo}(y_{t_0},t_0,p)\,,s)}{\partial x} f_{S|Y_{t_0+p}}\left(s\middle|y_{t_0}\right)\left(g(s,y_{t_0},p) - f_{S|Y_{t_0+p}}\left(s\middle|y_{t_0}\right)\right)ds > 0 \Leftrightarrow$$

$$\int_{s\in C} \frac{\partial w(x^{wo}(y_{t_0},t_0,p)\,,s)}{\partial x} g(s,y_{t_0},p)\,ds > \int_{s\in C} \frac{\partial w(x^{wo}(y_{t_0},t_0,p),s)}{\partial x} f_{S|Y_{t_0+p}}\left(s\middle|y_{t_0}\right)ds = 0$$

Thus the optimal choice when considering currency is higher than the one without considering it meaning that $x^{wo}\left(y_{t_0},t_0,p\right) < x^{w}(y_{t_0},t_0,p)$. This proves **(2), (3)** is proven analogously.